(51) International Patent Classification⁷: **C12P 21/00**	(74) Agents: **PARKER, Raymond, S. et al.**; Eli Lilly and Company, Lilly Corporate Center, Drop Code 1104, Indianapolis, IN 46285 (US).
(21) International Application Number: PCT/US00/28858	
(22) International Filing Date: 27 October 2000 (27.10.2000)	(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
(25) Filing Language: English	
(26) Publication Language: English	
(30) Priority Data: 60/165,153 12 November 1999 (12.11.1999) US	(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
(71) Applicant *(for all designated States except US)*: **ELI LILLY AND COMPANY** [US/US]; Lilly Corporate Center, Indianapolis, IN 46285 (US).	

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HARTMAN, Brian, Edward** [US/US]; 4444 North Washington Boulevard, Indianapolis, IN 46205 (US). **STROBEL, Robert, Joseph, Juni** [US/US]; 13807 Mill Stream Court, Carmel, IN 46032 (US). **SULLIVAN, Gary, Robert** [US/US]; 9022 Sargent Creek Drive, Indianapolis, IN 46256 (US).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: IMPROVED FERMENTATION PROCESS
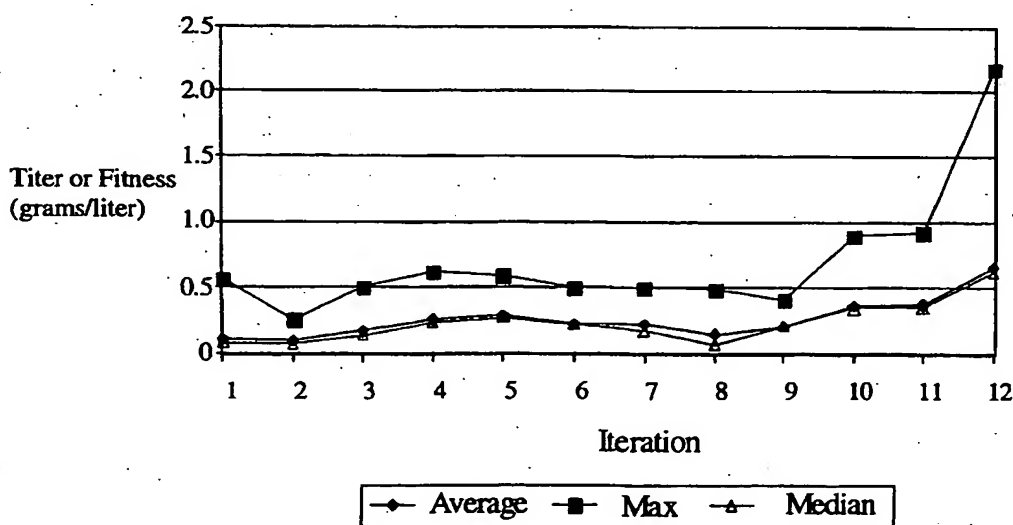
(57) Abstract: The present invention provides a process for optimizing a fermentation process for the production of a glycopeptide nucleus by means of particle swarm optimization. In addition, an improved fermentation medium for the production of the glycopeptide nucleus was found which is significantly different and more potent than media identified by more conventional optimization methods.

IMPROVED FERMENTATION PROCESS

FIELD OF THE INVENTION

The present invention relates to an improved fermentation process for the production of a glycopeptide nucleus from an *Amycolatopsis orientalis* or *Nocardia orientalis* medium based on particle swarm optimization.

BACKGROUND

Computational intelligence (CI) algorithms offer an alternative approach to *ad hoc* or statistically-designed optimization methods. The class of CI algorithms known as evolutionary algorithms(EA) seek to optimize an objective function by evolving a population of solutions in some manner. As their name suggests, biologically inspired operators such as crossover and mutation are used in combining good or "fit" solutions in order to improve the overall fitness of the population as an optimum is pursued. Stochastic elements of the search operators may lead to some portion of the population exploring very different, but feasible, regions of the search space. Solutions are not obtained, for example, by descending an objective function error gradient but rather by a pseudo-random search of the solution space. In general, no underlying assumptions about the objective function surface such as smoothness are necessary as they are with least-squares statistical design techniques.

The conceptual simplicity and ability to outperform classic methods in challenging real-world problems make CI algorithms attractive for study of fermentation media optimization. See e.g., Fogel D.B., "The Advantages of

Evolutionary Computation," In Proc. of BCE97: BioComputing

and Emergent Computation, Lundh D, Olsson B, Narayanan A

(eds.), Sinagapore:World Scientific, 1-11 (1997). However,

it must be fairly noted that a "typical" application of

5   these paradigms may involve very large population sizes

(i.e. number of candidate solutions studied at each

iteration) and may require many iterations to converge to an

acceptable solution. Given that fermentation medium design

can be expensive in terms of time, material, and labor

10   costs, these are two potential drawbacks when considering

the applicability of CI algorithms to medium design

problems. Fortunately, these issues have not precluded

research into the use of CI algorithms for this class of

problem. In a similar application, Weuster-Botz and Wandrey

15   used a GA to optimize fourteen medium ingredients for the

formate dehydrogenase fermentation, requiring only four

iterations. See, e.g., Weuster-Botz, D., et al., 1995.

"Medium Optimization by Genetic Algorithm for Continuous

Production of Formate Dehydrogenase," Process Biochemistry,

20   30(6), 563-571 (1995). Similarly, Backhouse, et al.

compared a GA with statistical techniques in the

optimization of four parameters in a yarn-spinning process

(five iterations), noting the importance of a good initial

randomization of the population. See, e.g., Backhouse,

25   P.G., et al., "A comparison of a genetic algorithm with an

experimental design technique in the optimization of a

production process," Journal of the Operational Research

Society, 48, 247-254 (1997).

To date, no one has attempted to optimize the

30   fermentation process for production of glycopeptide nuclei

using PSO. A variety of glycopeptide derivatives have been

shown to have significant activity as an antibacterial

agent. See, e.g., U.S. Patent Nos. 4,643,987; 4,698,327;
5,840,684; 5,843,437; and 5,843,889. Given the complexity
of components in a fermentation process, there is a need for
identifying optimized parameters for the production of, the
5   glycopeptide nucleus for use in synthesizing glycopeptide
antibacterial agents without having to conduct overwhelming
experimentation.

SUMMARY

10   The present invention provides a process for optimizing
a fermentation process for the production of a glycopeptide
nucleus (e.g., A82846B) from *Amycolatopsis orientalis* or
Nocardia orientalis (including mutants, variants or
recombinants thereof) comprising the step of determining key
15   component concentrations in the process by means of a
particle swarm optimization.

In another embodiment of the present invention, an
improved fermentation medium is provided wherein the
improvement is characterized by optimization of key
20   components of the medium by means of a particle swarm
optimization. The composition of the improved fermentation
medium as a result of applying the particle swarm
optimization is also provided.

In yet another embodiment of the present invention, a
25   fermentation medium for the production of a glycopeptide
nucleus from *Amycolatopsis orientalis* which contains no
animal source material (ASM), referred to herein as "ASM-
free" fermentation medium. The ASM-free medium comprises
cane molasses, hydrolyzed soybean flour and yeast. In a
30   preferred embodiment, the medium also includes corn gluten.

Definitions

As used herein, the term "particle swarm optimization" or "PSO" refers to a computational intelligence algorithm as described in Eberhart, R.C., et al. , "A new optimizer using particle swarm theory," In Proc. Sixth Intl. Symposium on Micro Machine and Human Science, (Nagoya, Japan), Piscataway, IEEE Service Center, 39-43 (1995); Kennedy, J., "The particle swarm: Social adaptation of knowledge," In Proc. 1997 IEEE International Conference on Evolutionary Computation (Indianapolis, IN), Piscataway:IEEE Service Center, 303-308 (1997), and Eberhart, R., et al., "Evolutionary Computation Implementations," Computational Intelligence PC Tools, Boston, AP Professional, 212-226 (1996), and are hereby incorporated herein by reference, to the extent that they explain, further enable, provide a basis for or describe the subject matter to which is referred to in the specification.

"Key components" refer to chemical ingredients of the fermentation process that have significant effects on the yield of the desired product (e.g., glycopeptide nucleus).

"Titer" refers to the standard of strength of a volumetric test solution, i.e., the assay value of an unknown measure by volumetric means.

DESCRIPTION OF THE FIGURES

Figure 1 represents the PSO Average and Maximum Titer by Iteration.

DETAILED DESCRIPTION

A significantly improved medium for the production of a glycopeptide nucleus from *Amycolatopsis orientalis* was found using particle swarm optimization. Eleven medium

ingredients were optimized using a total of 240 shake-flask
fermentations (12 iterations with a population size of 20;
each "individual", or "agent" in the population represents a
shake flask fermentation). The best medium improved
5   glycopeptide titer > 2-fold over *ad hoc* experimentation.
Whereas, a traditional statistical design approach using
similar numbers of shake-flasks improved the titer 40% by
optimizing five ingredients screened from the original 11.
Surprisingly, the PSO approach yielded a different optimal
10   medium than the medium derived through statistically
designed experimentation.

        PSO is a recent addition to the CI field. It is
similar to well known EAs such as genetic algorithms but is
defined in a social context as opposed to a biological
15   context. The individuals in the population retain memory of
known good solutions as they continue to search for better
solutions, unlike EAs where knowledge is destroyed between
generations. Interaction among parameters is thought to
enhance the ability of the PSO algorithm to find good
20   solutions. PSO seems to offer a powerful yet simple-to-
implement paradigm with a maximum of two algorithm
parameters that must be set prior to its use. This
simplicity makes the PSO algorithm appealing as a starting
point for forays into solving "real world" problems with CI
25   algorithms. Finally, concerning the algorithmic issues
above, PSO is reported to work well with relatively small
population sizes and to converge quickly over the range of
problems to which it has been applied to date (i.e.
determining artificial neural network weights to solve the
30   XOR problem. See, e.g., Eberhart, R.C., et al., "A new
optimizer using particle swarm theory," <u>In Proc. Sixth Intl.</u>
<u>Symposium on Micro Machine and Human Science</u>, (Nagoya,

Japan), Piscataway, IEEE Service Center, 39-43 (1995) and

Eberhart R., et al., "Evolutionary Computation

Implementations," Computational Intelligence PC Tools,

Boston, AP Professional, 212-226 (1996).

5   The usual method of optimizing early-phase fermentation

media is via the application of statistical design (SD)

methods, such as screening and response surface experiments.

These are inherently local optimization methods. PSO is a

global optimization method and takes advantage of the

10   intuitive fact that there are likely several 'optimal'

fermentation mediums which may be located in the

fermentation ingredient search space.

Applicants explored the applicability of PSO to

fermentation medium optimization for production of

15   glycopeptide nucleus in an attempt to identify advantages of

using an alternative approach to statistical design (SD)

methods. Classical ad hoc, or one-factor-at-a-time,

experimentation was also conducted for comparison purposes.

This work was done as an extension of efforts being made to

20   replace an animal-source nutrient (pork skin residue) in a

previously developed process for making the glycopeptide

nucleus.

Applicants have discovered that the use of PSO provides

several advantages, such as (i) lack of sensitivity to

25   (initial) ingredient ranges, (ii) ability to optimize in

higher dimensions, and (iii) global search vs. sequential

experimentation.

The range of each ingredient to be included in the

optimization study is typically set by the fermentation

30   scientist prior to designing the experiments. The

fermentation scientist balances knowledge of ingredient

ranges which have been successful in the past with the

desire to explore broader ranges, keeping in mind the need to minimize the number of experiments to keep labor and material costs low and to meet ever more aggressive product development time-lines. One advantage of SD is its ability to alert the experimenter that an optimum does not lie in the design space currently being studied and to indicate where the design space must be moved to pursue an optimum. But the need to run additional experiments is required, possibly delaying the project and adding additional costs. An optimization paradigm more robust to initial ingredient ranges could offer advantages from the standpoint of not requiring additional screening experiments.

As mentioned above, applying SD to a problem with more than four or five independent variables typically requires a preliminary screening design in order to reduce the number of independent variables and keep the factorial load to a manageable level. Often, the results of the screening experiments are straightforward. For example, in a fermentation medium optimization experiment, ingredients that have a significant negative linear effect with respect to the objective function (titer) may be effectively removed from further consideration. But often the analysis of screening experiment requires the aid of an experienced statistician. For example, does an ingredient with a slightly negative main effect still warrant inclusion in the experiments because it has a slightly positive interaction with another ingredient? Consider also the case where several ingredients exhibit similar effects on the objective function at the screening experiment phase or where higher order interactions are suspected. Successful resolution to the question of which ingredients to remove and which ingredients to keep for further study can be crucial to

7

success of the project. Either bypassing the screening stage, and thus avoiding any decisions regarding whether or not to keep an ingredient in the study, or screening out only major negative effects and keeping the remaining
5   ingredients in the plan could be advantageous.

SD techniques are sequential in nature - each stage of the experiment requires success in the prior stage(s). While conceptually simple, the proper application of SD methodologies can become technically demanding and require
10   the assistance of a professional statistician to ensure success. Often some subjective decisions may have to be made when planning the next stage of experiments depending on the strength of the results from the just completed experimental stage. The experimenter must balance the need
15   to "be bold" in planning further experiments with the constraint of operating in the current design space.

In a fermentation medium optimization problem with several ingredients it is not uncommon that more than one optimal solution exists - different mixtures of ingredients
20   may provide similar titer results. Optimizing globally via a "population of solutions" may provide the researcher with not only an optimized fermentation medium, but with several candidate media to explore as the individuals move throughout the search space. Sequential optimization
25   techniques like SD may also provide more than one candidate optimum, but these optima will tend to lie close together due to the local nature of the search process. The "richness" of solutions is another facet of the evolutionary algorithmic approach to be explored in this work. To
30   summarize, the intent was to apply a novel CI algorithm to a "standard" problem typically solved with SD approaches. The ability of PSO to simplify the optimization process with

respect to ingredient range selection, preclusion of any

screening stages, and the ability to globally optimize the

objective function in high dimensions, thus possibly

providing a "richer" set of solutions were of interest to

5   Applicants.

The following examples illustrate the optimization of

the fermentation process for the production of A82846B

glycopeptide nucleus from using PSO. The A82846B

glycopeptide nucleus has the following structure:

10
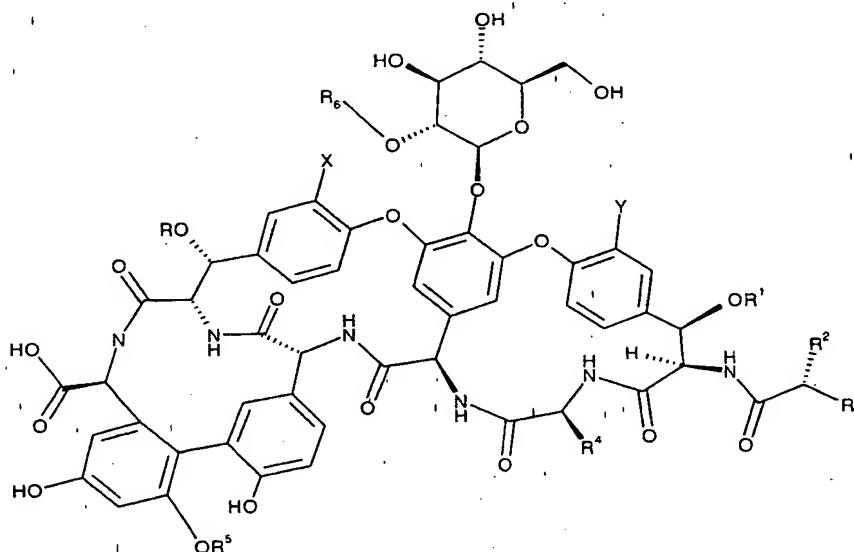
wherein R and R^6 are 4-*epi*-vancosaminyl, R^1 is hydrogen, R^2

is $NHCH_3$, R^3 is $CH_2CH(CH_3)_2$, R^4 is $CH_2(CO)NH_2$, R^5 is

hydrogen, and X and Y are Cl. The A82846B glycopeptide

nucleus may alternatively be produced from *Nocardia*

15   *orientalis* strains (e.g., NRRL 18098, NRRL 18099, and NRRL

18100) which is described in U.S. Patent No. 5,312,738,

incorporated herein by reference. Although the following

Examples illustrate the use of PSO to optimize the

fermentation media from a specific strain of *Amycolatopsis*

20   *orientalis*, those skilled in the art will appreciate that

other strains and/or mutants, variants or recombinants

thereof may be optimized using this process as well. For

example, one may use mutant strains that are produced using
Ultraviolet (UV) mutation procedures well known to those
skilled in the art. One may select a specific mutant based
on pigmentation which is also well know to those skilled in
5   the art.

EXAMPLES

Preparations

<u>*Glycopeptide Vegetative Flask Medium (GV-003):*</u>
10   The GV-003 broth was prepared by combining the
following components in the listed respective amounts.

Component	QA	Grams/Liter	% w/v
Glucose	QA020N	40	4
Phytone Peptone (BBL)	NA	15	1.5

No pH adjustment

15   <u>*Glycopeptide Bump Medium (GB-002):*</u>
        The GB-002 broth was prepared by adding the following
ingredients in the listed order. The pH was adjusted to 7.0
before addition of the calcium carbonate.

Component	QA/QD	Grams/Liter	%w/v
Glucose	QA020N	10	1
Corn Starch Powder	QA055L	5	0.5
Basic Yeast	QA220M	5	0.5
Hy-Soy™ (Quest)	QD463T	5	0.5
CaCO$_3$	QA209P	1	0.1

20
<u>*Fermentation Media:*</u>
        The producing culture, an UV mutated strain from
Amycolatopsis orientalis, was preserved and stored in liquid
nitrogen. Cultures were thawed and inoculated into GV-003

10

broth to initiate a seed train for the PSO studies. After
48 hours, a small amount (0.1 mL) of GV-003 culture was
transferred to second-stage seed broth, GB-002. After 48
hours growth in GB-002, 1.0 mL aliquots were transferred to

5   flasks containing 50 mL of media designed through the PSO
algorithm. All PSO media were buffered (pH 6.7) with 200 mM
filter-sterilized 3-(N-morpholino)propanesulfonic acid
(MOPS) added after flasks had been autoclaved (20 min.) and
cooled. PSO cultures were grown at 34.5°C for 6 days in a

10 shaker incubator set at 250 rpm with a relative humidity of
70%. Once a given set of shake-flask cultures was completed
(one iteration), samples were taken from the flasks and sent
to analytical laboratories for quantitative HPLC analysis of
glycopeptide factors. A complete fermentation cycle,

15 including preparation, inoculation, fermentation, assay, and
setting a culture for the next round of experiments,
required about 14 days.

PSO Methods

Basic algorithm:

20   Given a population of N individuals each individual i,
$i \in \{1,…,N\}$, has associated with it a position vector X_i on
D dimensions. The quantity $X_i[d]$, $d \in \{1,…,D\}$, represents
the dth parameter setting for individual i. The dimension D
is problem dependent and depends upon the number of free

25 parameters to be optimized in the system. At each iteration
k an individual's position vector $X_i(k)$ is updated by adding
a $\Delta X_i(k)$ vector, denoted V_i, to the current position. The
vector V_i represents an individual's velocity in the D-
dimensional problem hyperspace. Optimization is effected by

30 modifying the V_i vectors of all the individuals in the
population as shown below in equation 1.

Equation 1: Position update equation

$$X_i(k+1) = X_i(k) + V_i(k)$$

As the individuals, represented by particles, move throughout the problem space each retains knowledge of its best objective function value, denoted $pbest_i$ (for "personal best"), and the position in parameter hyperspace associated with that value of $pbest_i$, denoted $X_{pbest,i}$. The difference $X_{pbest,i} - X_i$ represents the geometric distance between an individual's current position and its best position found thus far. Adding a stochastic element in the form of a random positive number η results in an individual's update equation 2 given below.

Equation 2: Pbest velocity component update equation

$$V_i(k+1) = V_i(k) + \eta(X_{pbest,i} - X_i)$$

This update equation results in a stochastic tendency for each individual to return to its previous best position. This can be analogized with a human's tendency to remember and return to regions in the psychological space which have seemed beneficial or promising in the past.

To incorporate an interaction element the concept of a neighborhood is employed, defined as the nearest i - s and i + s individuals, for a neighborhood of size s, including the ith individual. In the context of this neighborhood construct, each individual retains knowledge of the overall best value of the objective function found thus far for the entire neighborhood, denoted gbest (for "global best"). The position in the problem space associated with the global best value gbest is denoted X_{gbest}. The vector $X_{gbest} - X_i$ represents the distance from individual i's current position and the position associated with the overall best objective

12

function value found thus far in the individual's
neighborhood. The interaction component of the algorithm
then consists of an update to V_i for each individual as
shown below in Equation 3.

5   <u>Equation 3</u>: Gbest velocity component update equation

$$V_i(k+1) = V_i(k) + \eta(X_{gbest,i} - X_i)$$

To prevent computer overflow as well as to define the
granularity of the search a parameter defined as V_{max},
representing maximum velocity, is defined and incorporated
10   into the algorithm. At each iteration the newly calculated
value of V_i is compared to V_{max} and the minimum value is
chosen for $V_i(k+1)$, preserving the sign of the current
direction. Combining (2) and (3) yields the complete
velocity update equation 4 below (note that the velocity may
15   be positive or negative).

<u>Equation 4</u>: Velocity update equations

$$V_i(k+1) = MIN\{V_i(k) + \eta_1(X_{pbest,i} - X_i) + \eta_2(X_{gbest,i} - X_i), V_{max}\}$$

Metaphorically, individuals (represented by particles)
20   are flown through D-dimensional hyperspace in search of good
solutions. Individuals are guided in their search by
knowledge of where the best solutions have been found thus
far. The preservation of information between iterations
sets the PSO algorithm apart from many evolutionary
25   algorithms where information is typically destroyed between
generations. (*Elitism*, or the introduction of one of the
most fit individuals in generation *k* into generation *k+1* is
an information preservation technique sometimes used in
evolutionary algorithms, though not usually on the scale of
30   PSO). The construct of the update equations ensures

13

individuals will 'fly by' known good regions and allocate some search time to unknown regions.'

The PSO algorithm, was coded as a MATLAB (The MathWorks, Cambridge, MA) m-file. The algorithm was written
5   to accept tab-delimited ASCII files containing information about the current iteration, such as coded ingredient levels, current ingredient delta values (e.g. agent velocities), values and location of each individual's best fitness (pbest values), the titer results of the current
10   fermentation (e.g. current fitness values), etc. New files are output containing the next set of experiments. In this manner the algorithm was run one iteration at a time, dependent upon the arrival of updated fitness information from the current set of experiments.

15   Microsoft Excel (Microsoft Corp., Redmond, WA) and JMP (The SAS Institute, Cary, NC) software were used to manipulate the various data files generated, for example de-scaling the coded ingredient levels returned by the PSO algorithm into the actual ingredient concentrations in
20   grams/liter. The various algorithmic details are described next.

Setting up the search space:

The number of ingredients as well as the allowable range of each ingredient defines the ingredient space in
25   which the PSO population will operate. The determination and application-related issues of each are discussed below.

Number of ingredients: The fermentation medium initially provided contained six ingredients, including a complex animal-source nutrient. Based on the composition of
30   the current medium, the desire to remove the animal-source

14

raw material, and overall production cost targets, the development scientist selected 11 ingredients for inclusion in the study.

<u>Ingredient ranges</u>: The development scientist, using

5   his judgement and prior experience, sets an initial range for each ingredient. These initial ranges were then linearly coded into [-10.0,10.0] for use in the PSO algorithm. For PSO, the determination of ingredient ranges is primarily to allow for initialization of the population

10   (see below) and to start the PSO population in a space where there is a reasonable chance of success.

Ingredient levels below 0.0 (uncoded) are infeasible. Thus, during the position update step of the algorithm, if any ingredient fell below -10.0 (coded), it was then clamped

15   to -10.0 (coded). A question arises as to what to do about the velocity value for an ingredient when it is clamped to its minimum value. Consider an ingredient that is heading towards its minimum value (implying negative velocity) and reaches the lower bound. If the individual attempted to

20   move to a different portion of the ingredient space including non-zero values of the ingredient, some time may be 'wasted' in that at least one iteration would likely be required to change the value of the velocity from negative to positive, after which the ingredient level could 'move

25   away' from 0.0. For this reason, any time an ingredient was clamped to its lower bound, the associated velocity was set to 0.0.

Conversely, the upper end of the ingredient ranges was set to 100.0 (coded), essentially providing for an unbounded

30   search of the ingredient space. The constraints above

result in the modifications to the position and velocity
update equations shown below.

Equation 5: Modified position update equation
$$X_i(k+1) = MAX\{X_i(k) + V_i(k), -10.0\}$$

Equation 6: Modified velocity update equation
$$V_i(k+1) = MIN\{V_i(k) + \eta_1(X_{pbest,i} - X_i) + \eta_2(X_{gbest,i} - X_i), V_{max}\}$$

If $X_i(k+1) \leq -10.0$ then $V_i(k+1) = 0.0$ and $X_i(k+1) = -10.0$

Setting Vmax:

Vmax was set to +/- 20% of the coded initial ingredient
ranges, or +/-2.0. This is commensurate with the typical
'rules of thumb' for using PSO and other computational
intelligence tools with a similar granularity construct.

Population size:

A population size of 20 was selected for this work. It
was felt this population size would allow for a complete
trial (consisting of 20 shake-flasks) to be prepared in 4-8
hours depending on the number of resources dedicated to the.
effort. Minimizing the amount of effort required for
preparing the PSO shake-flask media was considered important
in light of the high workload in the area.

Number of iterations:

It was hoped the project would run through 15
iterations of experiments. Although 15 iterations is
certainly a 'few' iterations in the context of applying
computational intelligence algorithms, it still represented
a fairly large amount of development time (15x20 = 300
planned experiments). One way to explore the reduction of

overall development time is to investigate the application
of PSO and SD methods in a hybrid sense. To address this
second project goal, the selection of a 'novel' location in
ingredient space would be done after the sixth iteration.

5   This would provide the PSO algorithm five iterations (in
addition to initialization) before the selection of a medium
was made.

Objective function:

Much discussion centered on how many project goals to
10   encode in the objective function. In addition to product
titer, broth viscosity was another response factor
considered for inclusion into the objective function.
Another metric considered was the percentage of the desired
chemical entity secreted by the *A. orientalis* production
15   culture, as at least two other chemical analogs were usually
produced in addition to the analog of interest. To meet
downstream purification targets, the desired chemical entity
had to represent a certain percentage (61%) of the overall
amount of the three analogs produced.

20   Several candidate objective functions were considered
to produce a figure of merit that would seek to reward
higher titer, lower viscosity, and meeting or exceeding the
percentage goal for the desired glycopeptide factor. The
possibility of simply having the scientist rank-order each
25   set of experiments was considered as well. In the end,
concerns about trying to achieve too many goals at once
during an initial foray into this type of project led the
team to decide to keep the objective function
straightforward, including only the potency metric. It
30   should be said that often media that provide for good titer
values usually have the other desirable characteristics as

well.   While serendipity may come into play with respect to
the achievement of these ancillary goals, they are also
somewhat interrelated.   For example, fermentation media
which become very viscous over the course of the
5   fermentation are less likely to promote good culture growth
and chemical synthesis as the high viscosity provides a less
than optimal environment due to impediments on mixing and
oxygen mass transfer.   Therefore, fermentation media that
provide for high culture growth and product synthesis
10   generally must usually exhibit favorable viscosity
characteristics in order to achieve the high growth and
synthesis in the first place.

Samples were taken twice during the course of a PSO
iteration to measure glycopeptide titer (potency), and it
15   was decided to scale the titer values.   The resulting
objective function used to score the fitness of the PSO
population after each trial was then:

Equation 7: Titer (Potency) Objective Function

$$\frac{Max(Potency - sample - 1, Potency - sample - 2)}{1000}$$

20   *Neighborhood size:*

Keeping with the goals of the experiment, it was
decided to use an initial neighborhood size of 3
individuals.   After five iterations, the neighborhood size
would be increased to 5 individuals.   The neighborhood size
25   would then be increased to its final value of 7 (or possibly
more) for the last five iterations. The basic intent was to
use a more 'localized' version of the algorithm for the
first part of the project, with the goal being to promote
search of several regions of the ingredient space at once.

As the algorithm progressed and good solutions were found in the various regions of ingredient space - the 'neighborhoods' - the algorithm would become more global in nature, with the best solution(s) having a greater impact on a larger portion of the PSO population.

Initialization of algorithm:

In CI applications it is common practice to initialize the population at random locations in the problem space, and the authors have used this approach in the past. For this work a different approach was used, following from the experience that CI algorithms typically require several iterations to become organized, after which real progress is made. The initialization of the population was thus effected in a manner similar to the generation of a screening design, with the intent to make the rows and columns of the design matrix as orthogonal and independent as possible. The design matrix used to initialize the PSO experimentation is provided in Table 1 below.

Table 1

Initial PSO design matrix (ingredients in grams per liter)

	Glucose	MgSO$_4$	(NH$_4$)$_2$SO$_4$	KH$_2$PO$_4$	KCl	Cane Molasses	Corn Gluten Meal	Hy-Soy™	Basic Yeast	Cotton Seed Flour	Corn Steep Powder
Flask 1	27.5	5.0	7.5	0.4	3.3	15.0	22.5	1.3	2.5	15.0	2.5
Flask 2	10.0	20.0	10.0	0.0	1.0	60.0	0.0	5.0	10.0	0.0	0.0
Flask 3	10.0	20.0	10.0	0.5	10.0	0.0	0.0	5.0	0.0	20.0	10.0
Flask 4	10.0	0.0	10.0	0.0	10.0	60.0	30.0	0.0	10.0	20.0	0.0
Flask 5	62.5	15.0	2.5	0.1	3.3	15.0	22.5	1.3	7.5	15.0	2.5
Flask 6	27.5	5.0	2.5	0.1	7.8	15.0	7.5	2.5	5.0	15.0	7.5
Flask 7	27.5	15.0	2.5	0.1	3.3	45.0	7.5	3.8	7.5	15.0	2.5
Flask 8	62.5	15.0	7.5	0.1	3.3	15.0	22.5	3.8	2.5	5.0	7.5
Flask 9	10.0	20.0	0.0	0.0	10.0	0.0	30.0	3.8	2.5	0.0	10.0
Flask 10	62.5	15.0	7.5	0.4	7.8	45.0	7.5	0.0	10.0	5.0	2.5
Flask 11	27.5	5.0	7.5	0.4	3.3	15.0	7.5	1.3	7.5	5.0	7.5
Flask 12	80.0	0.0	0.0	0.5	1.0	60.0	30.0	1.3	7.5	20.0	10.0
Flask 13	80.0	0.0	10.0	0.0	10.0	0.0	30.0	2.5	5.0	0.0	0.0

	Glucose	MgSO₄	(NH₄)₂SO₄	KH₂PO₄	KCl	Cane Molasses	Corn Gluten Meal	Hy-Soy™	Basic Yeast	Cotton Seed Flour	Corn Steep Powder
Flask 14	80.0	0.0	10.0	0.0	1.0	60.0	0.0	5.0	0.0	20.0	10.0
Flask 15	27.5	5.0	2.5	0.4	7.8	45.0	22.5	0.0	.0.0	5.0	2.5
Flask 16	10.0	20.0	0.0	0.5	1.0	60.0	30.0	3.8	2.5	0.0	10.0
Flask 17	80.0	20.0	0.0	0.5	10.0	0.0	0.0	0.0	0.0	20.0	0.0
Flask 18	80.0	0.0	0.0	0.5	1.0	0.0	0.0	0.0	0.0	0.0	0.0
Flask 19	45.0	10.0	5.0	0.3	5.5	30.0	15.0	5.0	10.0	10.0	5.0
Flask 20	62.5	5.0	2.5	0.1	7.8	45.0	7.5	2.5	5.0	5.0	7.5

Following the first experiment (above), the design matrix was reorganized in an attempt to take advantage of the neighborhood construct to be employed. A 20x11 matrix was constructed consisting of inter-column Euclidean distances. Using the first individual as a starting point, the individual with the least calculated distance from the first was re-indexed to be the second individual in the population. Next, the individual the least distance from the second individual was moved and re-indexed as the third individual, and so on. In this manner each neighborhood consisted initially of individuals with the least distance from each other in ingredient space. Table 2 reflects the updated distance-based design matrix.

Table 2

PSO position initialization (ingredients in grams per liter)

	Glucose	MgSO₄	(NH₄)₂SO₄	KH₂PO₄	KCl	Cane Molasses	Corn Gluten Meal	Hy-Soy™	Basic Yeast	Cotton Seed Flour	Corn Steep Powder
Flask 1	27.5	5.0	7.5	0.4	3.3	15.0	22.5	1.3	2.5	15.0	2.5
Flask 2	27.5	5.0	2.5	0.1	7.8	15.0	7.5	2.5	5.0	15.0	7.5
Flask 3	27.5	5.0	7.5	0.4	3.3	15.0	7.5	1.3	7.5	5.0	7.5
Flask 4	45.0	10.0	5.0	0.3	5.5	30.0	15.0	5.0	10.0	10.0	5.0
Flask 5	27.5	15.0	2.5	0.1	3.3	45.0	7.5	3.8	7.5	15.0	2.5
Flask 6	27.5	5.0	2.5	0.4	7.8	45.0	22.5	0.0	0.0	5.0	2.5
Flask 7	10.0	20.0	0.0	0.5	1.0	60.0	30.0	3.8	2.5	0.0	10.0
Flask 8	10.0	0.0	10.0	0.0	10.0	60.0	30.0	0.0	10.0	20.0	0.0
Flask 9	10.0	20.0	10.0	0.0	1.0	60.0	0.0	5.0	10.0	0.0	0.0
Flask 10	62.5	5.0	2.5	0.1	7.8	45.0	7.5	2.5	5.0	5.0	7.5
Flask 11	62.5	15.0	7.5	0.4	7.8	45.0	7.5	0.0	10.0	5.0	2.5

	Glucose	MgSO₄	(NH₄)₂SO₄	KH₂PO₄	KCl	Cane Molasses	Corn Gluten Meal	Hy-Soy™	Basic Yeast	Cotton Seed Flour	Corn Steep Powder
Flask 12	62.5	15.0	7.5	0.1	3.3	15.0	22.5	3.8	2.5	5.0	7.5
Flask 13	62.5	15.0	2.5	0.1	3.3	15.0	22.5	1.3	7.5	15.0	2.5
Flask 14	80.0	0.0	10.0	0.0	10.0	0.0	30.0	2.5	5.0	0.0	0.0
Flask 15	80.0	0.0	0.0	0.5	1.0	0.0	0.0	0.0	0.0	0.0	0.0
Flask 16	80.0	20.0	0.0	0.5	10.0	0.0	0.0	0.0	0.0	20.0	0.0
Flask 17	80.0	0.0	10.0	0.0	1.0	60.0	0.0	5.0	0.0	20.0	10.0
Flask 18	80.0	0.0	0.0	0.5	1.0	60.0	30.0	1.3	7.5	20.0	10.0
Flask 19	10.0	20.0	0.0	0.0	10.0	0.0	30.0	3.8	2.5	0.0	10.0
Flask 20	10.0	20.0	10.0	0.5	10.0	0.0	0.0	5.0	0.0	20.0	10.0

The velocity, or change in position, for each ingredient in each flask must also be initialized, and again it is common practice to initialize the PSO velocities in some random manner. For this work it was considered to initialize all the velocities to 0.0, again in the context of trying to give the PSO population the most efficient starting point and (lack of) initial trajectory. It was thought that perhaps a random initialization of the velocities might contribute to the PSO population requiring a several iterations simply to organize, such as when the combined velocity vector led an agent to a worse-performing region of ingredient space, from which it would seek to return. However, Applicants eventually decided to initialize the velocities to $N(0, V_{max})$ random values. Table 3 contains the initial velocity settings for all agents.

Table 3

PSO velocity initialization (coded values)

	Glucose	MgSO₄	(NH₄)₂SO₄	KH₂PO₄	KCl	Cane Molasses	Corn Gluten Meal	Hy-Soy™	Basic Yeast	Cotton Seed Flour	Corn Steep Powder
Flask 1	0.23	1.75	-1.61	0.28	1.48	1.85	0.00	-1.72	0.10	-1.62	-1.33
Flask 2	-0.49	-0.16	-1.46	-0.37	-0.26	-0.64	-1.11	-0.46	1.62	1.77	-1.09
Flask 3	-1.30	1.87	-0.80	-1.31	1.73	-0.11	1.76	1.83	1.27	1.50	1.58
Flask 4	-1.72	-1.21	0.31	-0.77	-1.96	1.64	-1.97	-0.92	1.39	1.20	0.67
Flask 5	-0.72	-0.12	1.72	1.21	0.63	-0.06	0.90	-0.59	1.81	1.94	-1.35
Flask 6	1.29	-1.04	0.62	1.00	-0.92	0.76	1.72	0.41	-0.34	1.47	-1.44

	Glucose	MgSO₄	(NH₄)₂SO₄	KH₂PO₄	KCl	Cane Molasses	Corn Gluten Meal	Hy-Soy™	Basic Yeast	Cotton Seed Flour	Corn Steep Powder
Flask 7	-1.65	1.65	-1.28	0.43	-0.15	0.57	1.40	1.08	1.49	0.22	0.77
Flask 8	-1.67	0.76	1.05	1.84	1.17	-0.67	0.67	1.26	-1.83	0.76	-0.29
Flask 9	0.59	1.37	-0.49	1.78	1.78	0.01	-0.21	0.29	-1.47	0.75	-1.00
Flask 10	1.46	0.26	0.34	0.39	-1.80	1.88	-0.29	-1.03	-0.97	-0.01	0.78
Flask 11	-1.64	-1.11	0.07	-0.08	0.49	-0.08	0.08	-0.73	-0.04	0.12	-1.99
Flask 12	-0.68	0.87	-1.33	-0.75	1.23	-1.36	0.00	-0.96	-1.57	-1.59	1.88
Flask 13	-1.02	-1.53	-0.11	-1.08	0.99	1.54	1.97	0.40	-1.33	0.55	1.18
Flask 14	-1.31	-0.23	-1.10	1.66	0.00	-0.67	0.16	0.95	0.78	0.24	1.89
Flask 15	-1.00	-0.08	-1.17	0.16	-0.07	-0.09	1.29	0.85	0.29	1.75	-0.67
Flask 16	-1.85	0.63	-0.81	-0.42	-0.14	-0.89	1.44	-0.85	-0.77	-1.78	0.30
Flask 17	0.02	0.75	-1.33	-1.72	0.30	-1.72	1.40	0.16	0.44	0.76	1.77
Flask 18	0.11	0.22	0.71	0.78	-0.52	-0.90	-0.55	0.48	1.82	-1.37	0.67
Flask 19	-1.95	-1.81	-0.66	-1.85	-1.71	-1.59	0.08	1.35	-1.34	-0.83	1.05
Flask 20	1.68	1.72	1.72	-1.12	-0.42	-0.08	-1.70	0.19	-0.70	-0.77	-1.76

Velocity damping:

Finally, one additional modification was made to the velocity update equation. The velocity is multiplied by a weight, or damping factor, which declines linearly over several iterations. For this work, the weight was initialized to 0.9 and the intent was to reduce the weight to 0.4 over 15 iterations. Similar to the momentum term in standard back-propagation of error optimization algorithms, the weight term works to 'smooth' movement in ingredient-space which could be helpful in finding any optima once promising regions are located. Considering this weight term, the final velocity update equation results, as given below.

Equation 8: Final velocity update equation

$$V_i(k+1) = MIN\{w[V_i(k) + \eta_1(X_{pbest,i} - X_i) + \eta_2(X_{gbest,i} - X_i)], V_{max}\}$$

If $X_i(k+1) \leq -10.0$ then $V_i(k+1) = 0.0$ and $X_i(k+1) = -10.0$

The results of applying PSO to glycopeptide optimization fermentation medium are described below. The results of maximizing potency are presented first and

22

compared to the results achieved in the parallel SD optimization work.

It should be noted that in the middle of the experiment it was determined that a key ingredient, $(NH_4)_2SO_4$, had become contaminated and had likely adversely affected the results of several of the PSO iterations. This problem remained unnoticed for several iterations of the PSO algorithm. However, the excellent results obtained using PSO suggest the algorithm was able to easily recover once the ingredient problem was corrected.

The total number of PSO iterations were reduced to 12 rather than the planned 15. The progression of fitness values by iteration is illustrated in Table 4 below and Figure 1.

Table 4

PSO medium optimization results for each iteration

(g of nucleus/liter)

	1	2	3	4	5	6	7	8	9	10	11	12
Flask 1	0.06	0.07	0.12	0.14	0.28	0.29	0.12	0.19	0.09	0.13	0.37	0.30
Flask 2	0.02	0.23	0.10	0.12	0.27	0.22	0.17	0.41	0.20	0.24	0.23	0.34
Flask 3	0.00	0.15	0.26	0.11	0.21	0.07	0.18	0.04	0.11	0.13	0.20	0.20
Flask 4	0.15	0.10	0.11	0.21	0.21	0.05	0.23	0.00	0.32	0.30	0.47	0.59
Flask 5	0.11	0.03	0.07	0.12	0.33	0.31	0.29	0.05	0.36	0.39	0.49	0.52
Flask 6	0.00	0.09	0.49	0.24	0.32	0.25	0.49	0.00	0.32	0.24	0.92	0.77
Flask 7	0.00	0.00	0.19	0.19	0.27	0.01	0.40	0.23	0.34	0.64	0.69	0.94
Flask 8	0.00	0.01	0.00	0.10	0.00	0.06	0.01	0.00	0.34	0.46	0.55	0.71
Flask 9	0.16	0.18	0.21	0.61	0.59	0.30	0.41	0.27	0.20	0.50	0.50	0.74
Flask 10	0.13	0.08	0.01	0.28	0.25	0.50	0.26	0.35	0.41	0.89	0.51	0.74
Flask 11	0.56	0.14	0.01	0.44	0.45	0.21	0.22	0.49	0.21	0.32	0.41	0.67
Flask 12	0.21	0.04	0.45	0.29	0.44	0.07	0.16	0.03	0.03	0.26	0.22	0.57
Flask 13	0.02	0.04	0.04	0.44	0.10	0.10	0.15	0.04	0.02	0.12	0.05	0.22
Flask 14	0.14	0.07	0.40	0.16	0.05	0.05	0.04	0.07	0.21	0.24	0.33	2.18
Flask 15	0.04	0.07	0.13	0.42	0.40	0.45	0.12	0.04	0.41	0.49	0.11	0.79
Flask 16	0.09	0.10	0.18	0.30	0.35	0.22	0.17	0.15	0.11	0.42	0.20	0.71
Flask 17	0.06	0.01	0.00	0.00	0.20	0.39	0.16	0.03	0.04	0.36	0.27	0.64
Flask 18	0.16	0.00	0.00	0.00	0.17	0.12	0.08	0.04	0.05	0.13	0.11	0.54
Flask 19	0.02	0.25	0.21	0.31	0.43	0.44	0.44	0.35	0.19	0.52	0.44	0.51
Flask 20	0.14	0.07	0.24	0.50	0.23	0.28	0.15	0.15	0.06	0.42	0.29	0.36
Average	0.10	0.09	0.16	0.25	0.28	0.22	0.21	0.14	0.20	0.36	0.37	0.65

| Median | 0.07 | 0.07 | 0.12 | 0.23 | 0.27 | 0.22 | 0.17 | 0.06 | 0.20 | 0.34 | 0.35 | 0.61 |
| Maximum | 0.56 | 0.25 | 0.49 | 0.61 | 0.59 | 0.50 | 0.49 | 0.49 | 0.41 | 0.89 | 0.92 | 2.18 |

The period of ingredient contamination from iteration 5 through iteration 9 can be readily seen as the average population fitness decreases slightly during this period. Once the situation was corrected, the average and maximum fitness values increased dramatically. Average fitness increased from 0.1 g/L to 0.65 g/L, or 550%. Similarly, median fitness increased from 0.07 g/L to 0.61 g/L, or 771%. Discounting the highest fitness result obtained in the 12th (final) iteration, the mean and median fitness values increased 270% and 250%, respectively. The medium achieving the best fitness (2.18 g/L) in the final iteration was comprised of 10 of the 11 possible ingredients, with 3 of the ingredients present in very minute quantities. In all, 12 x 20 = 240 individual flask trials were performed, requiring approximately 7 months to complete. By these measures, the results achieved using PSO were excellent.

In comparison, the SD project work yielded a medium with a fitness of 1.05 g/L, requiring 7 ingredients. A response surface design suggested that a simpler medium, consisting of only 3 of the 7 ingredients, would achieve a predicted potency of 1.1 g/L. In all, approximately 100 SD shake-flask trials were performed, requiring approximately 4 months to complete.

Beyond the level of fitness achieved, the difference in composition of the optimal fermentation media obtained is of interest and useful for comparison. Table 5 lists the composition of the best PSO medium along with the best SD-derived fermentation medium and the best medium obtained by the *ad hoc* approach. The titers shown are averages for

24

several flask trials (replicates). The fact that the PSO and SD-derived media are very different is readily seen, although some similarities are evident as well. It is interesting to note the best PSO result was obtained in the ingredient space defined by the initial PSO ingredient ranges. In contrast, the design space moved under the SD approach, with Basic Yeast at 12.5 g/L in the best SD results. Examination of the PSO population locations over all the experiments (not shown) indicates several ingredients were tested outside the initial ranges, but the overall tendency was to search within the pre-defined ingredient space as that was where the best solutions were located by the PSO population.

Table 5

Comparison of optimized media[1]

	PSO Medium	SD Medium	*Ad Hoc* Medium	*ASM* Medium
Ingredient (g/liter)				
Meat Peptone	0.0	0.0	0.0	22.5
Glucose	80.0	45.0	45.0	45.0
$MgSO_4$	0.0	1.2	1.2	1.2
$(NH_4)_2SO_4$	7.3	4.13	4.13	4.13
KH_2PO_4	0.15	0.16	0.16	0.16
KCl	8.6	10.0	5.5	5.5
Cane Molasses	2.6	10.0	0.0	0.0
Corn Gluten	23.1	0.0	20	0.0
Hy-Soy™	2.5	2.5	0.0	0.0
Basic Yeast	4.9	12.5	0.0	0.0
Titer (g/L)	**1.8**	**1.05**	**0.74**	**1.0**
Std Deviation	**0.09**	**0.05**	**0.06**	**0.05**
No. of Replicates	**4**	**4**	**4**	**4**

[1]All media contained $CaCO_3$ at 5.2 grams/liter

In examining the above results, it is clear the PSO algorithm was able to optimize the fermentation medium for the modified glycopeptide process. In addition, the PSO algorithm was shown to be robust in the face of some challenging issues. For example, the incidence of an

ingredient becoming contaminated had no more a deleterious
effect than the likely addition of iterations required to
reach high glycopeptide titer. Although it is only
conjecture, considering that iterations at risk consisted of

5   80-100 experiments, it is possible similar results might
have been realized in 140-160 PSO experiments, or 7 to 8
trials, which is very similar to the number required for SD
optimization. Furthermore, the ability to simply 'keep
going', rather than having to 'go back' and repeat one or

10   more experimental stages - as would likely be required in SD
- is certainly of value.

This result represents a very exciting discovery for
the area, not only with respect to the ability of a tool
like PSO to provide an optimized fermentation medium, but

15   also in terms of the value delivered to the project by
providing a fermentation medium that supports high product
titer. In addition, the optimized medium identified
enhances the prospects for commercialization of a
glycopeptide antibacterial agent by providing cost savings

20   as a result of increased productivity of the medium.

In addition, a fermentation medium was identified which
was free from animal source materials (ASM) without
sacrificing titer. In fact, the PSO medium using a
combination of cane molasses (or blackstrap molasses), corn

25   gluten, Hy-Soy™ (hydrolyzed soybean flour available from
Quest International, Norwich, NY) and yeast (e.g. Red Star
basic yeast, available from Red Star, Milwaukee, WI) as a
nutrient source provided a titer which was 180% higher than
a comparative medium using a meat peptone as a nutrient

30   source. The SD medium using cane molasses, Hy-Soy™ and
basic yeast as a nutrient source provided a titer which was

nearly equivalent to the comparative medium using meat peptone (105%). Therefore, both the PSO and SD ASM-free media provide viable alternatives for media containing animal-source materials.

5   Other useful alternatives for cane molasses include any sucrose related materials. Useful alternatives for Hy-Soy™ include other hydrolyzed soybean flours which may be hydrolyzed enzymatically, or by means of an acid or base hydrolysis. Suitable yeasts include any spray dried whole

10   yeast materials or yeast extracts. Suitable corn glutens include whole or hydrolyzed glutens and equivalents thereof.

CLAIMS

We Claim:

   1.   A process for optimizing a fermentation process
for the production of a glycopeptide nucleus from
Amycolatopsis orientalis or *Nocardia orientalis* comprising
the step of determining key component concentrations in said
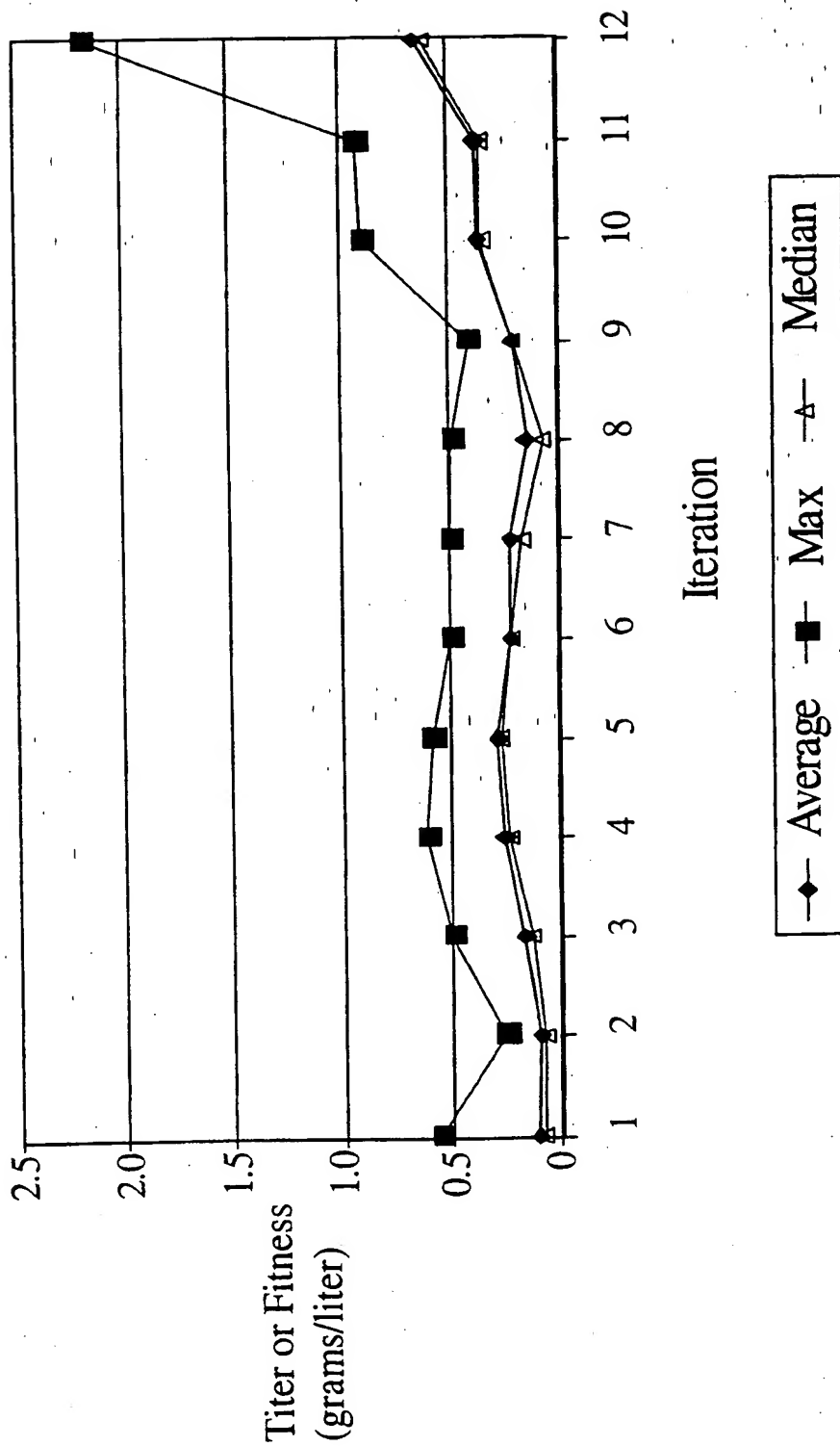process by means of a particle swarm optimization.

   2.   The process of Claim 1 wherein said glycopeptide
nucleus is A82846B.

   3.   An improved fermentation medium for the production
of a glycopeptide nucleus from *Amycolatopsis orientalis* or
Nocardia orientalis wherein said improvement is
characterized by optimization of key components of said
medium by means of a particle swarm optimization.

   4.   The medium of Claim 3 wherein said glycopeptide
nucleus is A82846B.

   5.   An ASM-free fermentation medium for the production
of a glycopeptide nucleus from *Amycolatopsis orientalis*
comprising cane molasses, hydrolyzed soybean flour and
yeast.

   6.   The fermentation medium of Claim 5 further
comprising corn gluten.

FIG. 1